ISMIR 2015 MALAGA
16th International Society for Music Information Retrieval Conference - October 26-30, 2015 Málaga, Spain

Score Following for Piano Performances with Sustain-pedal Effects

Bochen Li, Zhiyao Duan
Department of Electrical and Computer Engineering, University of Rochester

Summary

• Sustain pedal is commonly used in piano music after romantic era. (Fig.1)

• With sustain pedal pressed, sound is longer than the notated length. Sustained sound may overlap with next sound and confuses the system.

• Proposed spectral peak removal operation will remove partials of sustained sound and encourage the system to find correct positions.

• Evaluations on 50 randomly selected pieces from MAPS dataset show significant improvements on both accuracy and robustness.

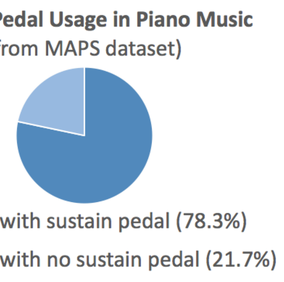


Figure.1

Piano Sound Properties

Without sustain pedal:

Pressing and holding the key will yield an impulse-like articulation, and releasing the key will let the damper touch the string quickly, thus the sound ceases.

With sustain pedal:

All dampers of all keys will never touch the string no matter if a key is pressed or not. Played sound will be sustained until string vibration ceases naturally. When the sustained sound spreads to the next onset, the mixture of energy will cause potential mismatch between audio and score, as Fig.2 shows

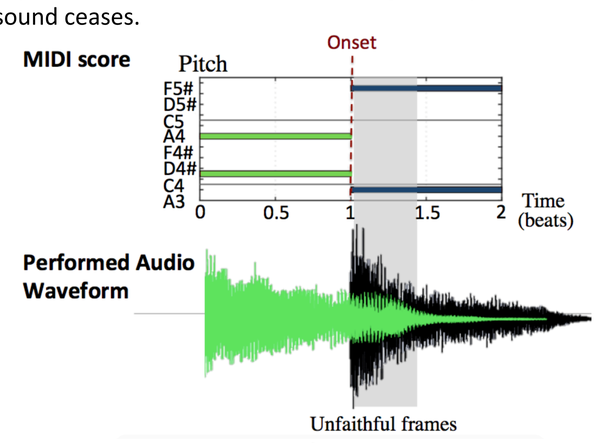


Figure.2

Method

1. System Framework

The system is based on the state space model proposed in [1]. The goal of the system is to infer the score state $\mathbf{s}_n$ (current position and tempo) from current and previous audio observations $y_1, \cdots, y_n$. This is formulated as an online inference problem of hidden states of a hidden Markov process, as fig.2 shows.

Observation model evaluates the match between an audio frame and the hypothesized state on the pitch content. This is calculated using the Multi-Pitch Estimation (MPE) likelihood model proposed in [2], which evaluates the likelihood of a hypothesized pitch set in explaining the magnitude spectrum of an audio frame.

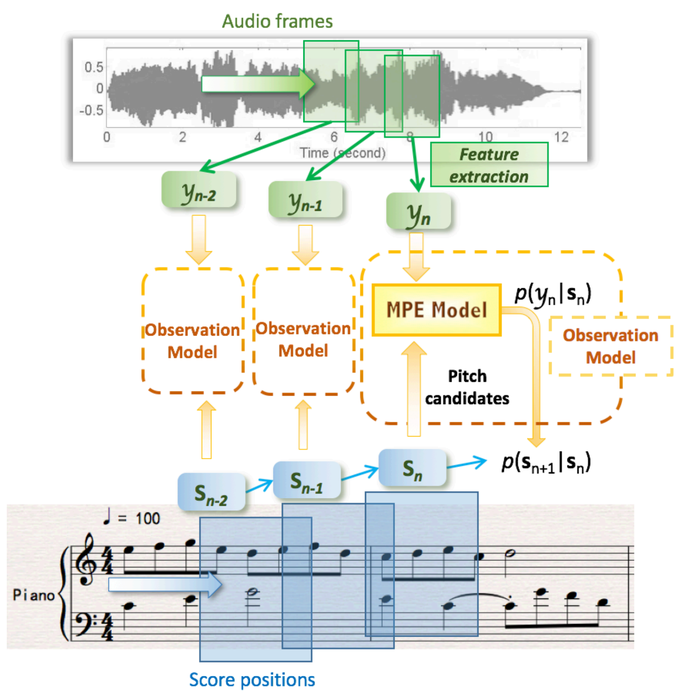


Figure.3

For piano music played with sustain pedal, frames within a period after a detected onset are potential unfaithful frames due to the sustain-pedal effects. We apply a spectra-based onset detection method in **feature extraction** part to locate unfaithful frames and remove spectral peaks in these frames from sustained sound. Fig.4 shows a detailed feature extraction flowchart.

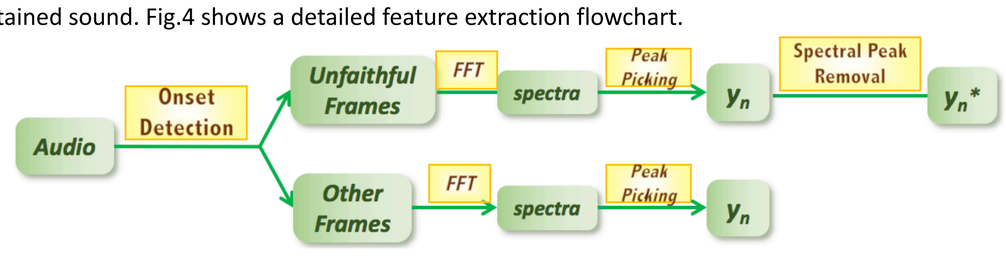
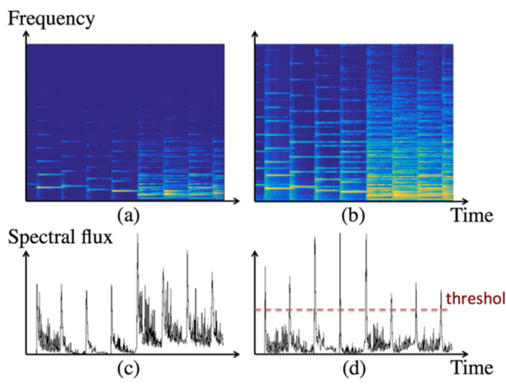


Figure.4

2. Spectra-based Onset Detection Method



(a). Spectrogram from STFT

(b). Logarithmic compression on spectrogram

$$\tilde{\mathbf{Y}}(n,k) = \log\left(1 + \gamma \cdot \mathbf{Y}(n,k)\right)$$

(c). summing positive temporal differences

$$\Delta_{\mathbf{Y}}(n) = \sum_k \left|\tilde{\mathbf{Y}}(n,k) - \tilde{\mathbf{Y}}(n-1,k)\right|_{\geq 0}$$

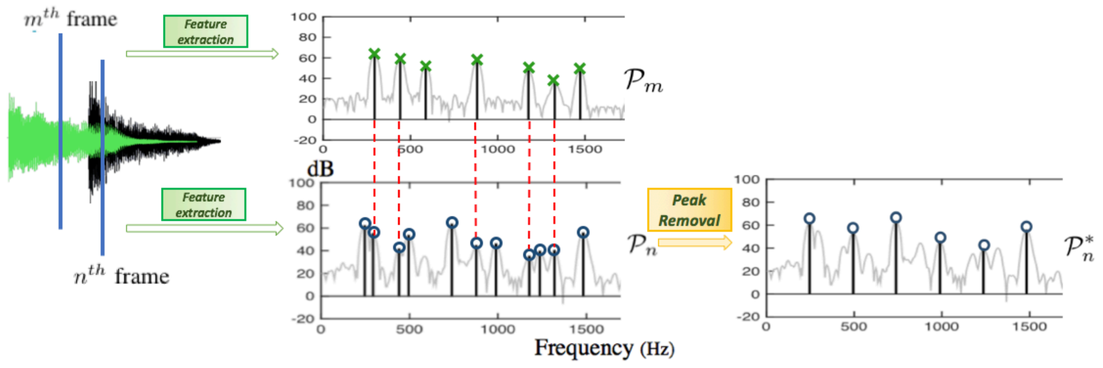
(d). normalize by audio energy

$$\tilde{\Delta}_{\mathbf{Y}}(n) = \Delta_{\mathbf{Y}}(n)/E(n)$$

Figure.5

3. Peak Removal Criteria

A peak in the n-th frame { ①whose frequency is very close to / ②whose amplitude is smaller than } those of a peak in the m-th frame



$$\mathcal{P}_n^* = \mathcal{P}_n - \left\{\langle f_i^n, a_i^n\rangle : \exists j \text{ s.t. } |f_i^n - f_j^m| < d, a_i^n < a_j^m\right\}$$

Figure.6

New Mismatch Introduced by Peak Removal

Although peak removal operation removes notes extended by the sustain pedal in the audio representation, in some case, it also removes notes that should remain according to the score, e.g, note D4 in Fig.7-a.

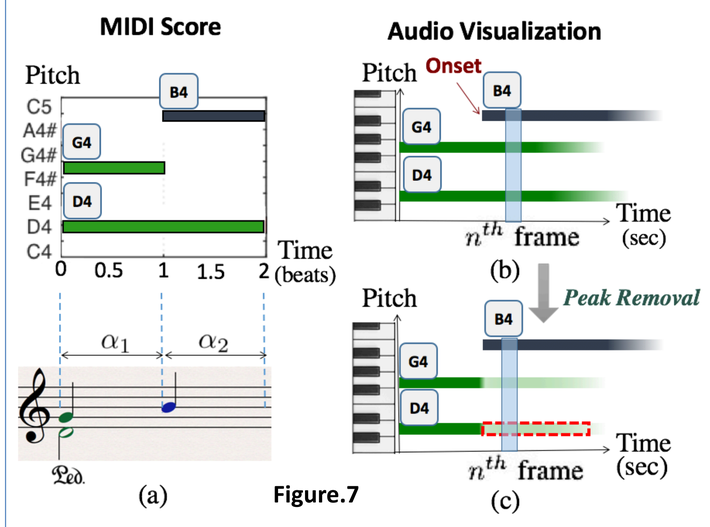


Figure.7

In this case, given n-th audio frame and 2 score position candidates α_1 and α_2, the true score position should be in α_2.

Recall **Observation Model:** $p(y_n|s_n)$
The system outputs estimated position by comparing $p(y_n|\alpha_1)$ & $p(y_n|\alpha_2)$

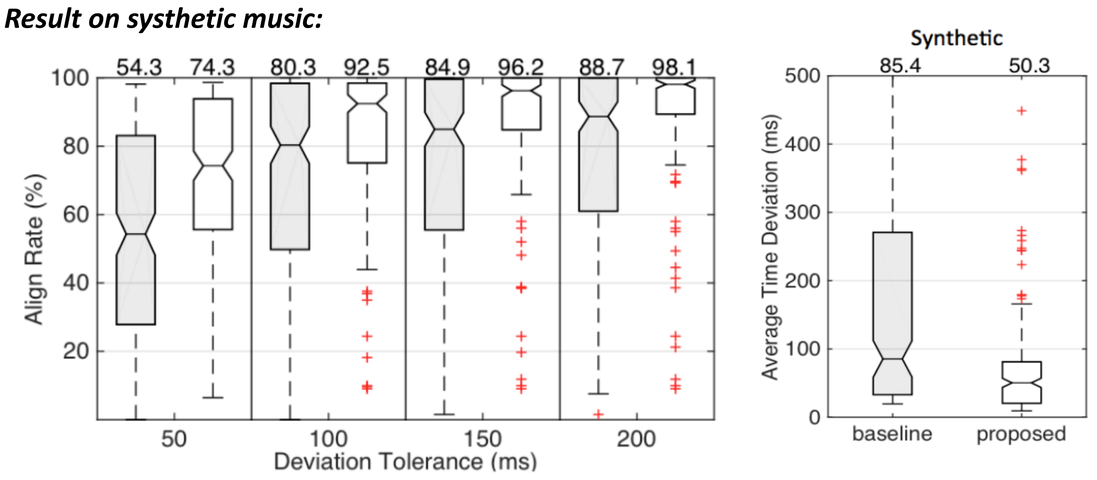
Before ***Peak Removal*** (Fig.7-b):
$p(y_n|\alpha_1)$: with support from D4, G4
$p(y_n|\alpha_2)$: with support from D4, B4
Hard to compare!

After ***Peak Removal*** (Fig.7-c):
$p(y_n|\alpha_1)$: none support
$p(y_n|\alpha_2)$: with support from B4
then $p(y_n|\alpha_1) < p(y_n|\alpha_2)$
Output α_2***!***

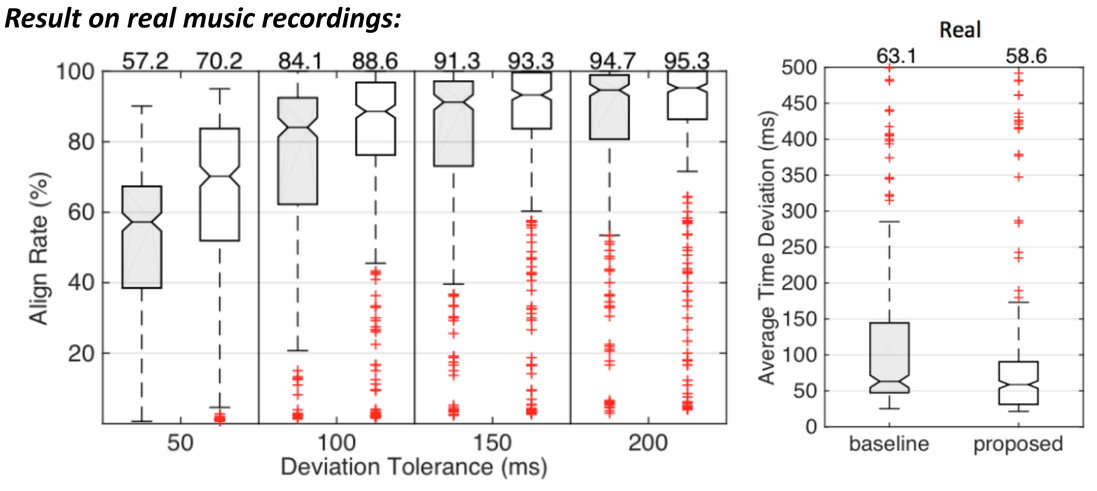
Conclusion: Peak removal operation reduces the mismatch caused by the sustain-pedal effects at the expense of introducing potential new mismatch caused by the removal of notes whose keys have not been released. However, this operation still helps the system to favor the correct score position even in this case

Experiments

Dataset: 50 music pieces from MAPS dataset (25 synthetic, 25 recorded)
Measurement: *Average Time Deviation (ATD), Align Rate (AR)*
Result on systhetic music:



Result on real music recordings:



References

[1] Zhiyao Duan and Bryan Pardo, **A state space model for online polyphonic audio-score alignment**, in Proc. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.

[2] Zhiyao Duan, Bryan Pardo, and Changshui Zhang, **Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions**, *IEEE Trans. Audio Speech Language Process.*, vol. 18, no. 8, 2010.